c SOUTHEAST ASIAN LINGUISTICS SOCIETY SEALS

Journal of the Southeast Asian Linguistics Society
JSEALS 12.2 (2019): i-viii
ISSN: 1836-6821, DOI: http://hdl.handle.net/10524/52453
University of Hawai'i Press

MUAK SA-AAK: CHALLENGES OF AN EXTENSIVE PHONEME INVENTORY FOR A CONTAINED LATIN-BASED ORTHOGRAPHY

Elizabeth Hall
Payap University Linguistics Institute
ellie_hall@sil.org

Abstract
Many Asian languages have extensive phoneme inventories and distinguish suprasegmental features like tone, voice quality, or vowel length. These features are well represented in extensive alphasyllabaries such as Burmese, Khmer or Thai. Apart from challenges of socio- or psycholinguistic nature regarding the need for diacritics, digraphs, or special characters when the language needs more symbols than are available in the contained Latin alphabet, considerable technical challenges may result when new orthographies for minority languages in Asia are based on the Latin alphabet. Several of these challenges were faced when developing a Latin-based orthography for Muak Sa-aak, a Palaungic tonal language of Shan State, Myanmar. This paper will give an overview of the design and features of Latin-based Muak Sa-aak orthography and illustrate computer-technical implications of the choices made.[1]

Keywords: Muak Sa-aak, Palaungic, orthography design, tone marks, text-processing
ISO 639-3 codes: ukk

1 Introduction

This study presents Muak Sa-aak phonemes and their orthographic rendering, followed by a short evaluation of the orthography and a discussion of the resulting technological challenges in regard to Unicode and use of the orthography in computer applications. Considerable technical challenges arise in orthography design for minority languages in Mainland Southeast Asia. Fonts and keyboarding solutions may be required, even if the orthography is based on the national or a majority language writing system. Many Southeast Asian alphasyllabaries such as the one for Tai Khuen (Owen 2017) or Shan (Inglis 2017) include characters that change shape according to the letters that appear with them. Even though 'smart fonts' can choose the correct shape according to the environment, software must support the rendering technology required for these fonts. Furthermore, since many minority languages based on non-Latin scripts are not specifically supported by major software vendors, even when special fonts and keyboard layouts are produced, software updates may result in incompatibilities so that suddenly a keyboard or font no longer works.

Due to the extensive segmental and suprasegmental phoneme inventories typical for Southeast Asian languages, Latin-based minority language orthography development may especially face challenges regarding the need for special characters not found in the contained English alphabet, again resulting in a need for special fonts and keyboards. Attempts to add these characters by use of other characters on an English keyboard, such as the use of numerals, punctuation or symbols as letters, do not conform to Unicode Standards. Other challenges in orthography development include spellchecking and autocorrect, language tagging of text, sort orders, capitalization, word selection, and hyphenation, to mention just a selection of what should be considered in orthography development (cf. Hosken, 2003; SIL, 2018). To illustrate some of

[1] An earlier version of this paper, *An Austroasiatic tonal language: Technical challenges of its Roman-based script* was presented at the 26th Southeast Asian Linguistics Conference, Manila, Philippines, 26-28 May 2016.

those issues, this paper will discuss the Latin-based orthography of Muak Sa-aak, a minority language spoken in Myanmar, developed in 2011. Adjustments include representation of tone, large vowel inventories, and reduced syllables. The resulting technical challenges are discussed especially in light of compliance with the Unicode standard.

Muak Sa-aak is an Austroasiatic language, belonging to the Angkuic branch of the Eastern Palaungic subgroup (Hall 2013). The Muak Sa-aak are located mainly in eastern Shan State of Myanmar, with an estimated population of 4,460. Four villages are also known to be located in China. In the Muak Sa-aak area, various Burmese-based alphasyllabaries are in use, although some minority groups use Latin-based alphabets. Since Muak Sa-aak never previously had an orthography, an initial orthography workshop was held with three Muak Sa-aak native speakers in Thailand using modified participatory methods as described by Page (2013). Various speakers had stated a preference for a Latin-based orthography; it was chosen for aesthetic and sociolinguistic reasons, such as orienting themselves toward the Latin-based writing conventions of a politically powerful minority group in their area. Adjusting the contained (cf. Nag 2007) Latin alphabet for a tonal language with a large vowel inventory presents difficulties because of the limited grapheme inventory, while the benefit in comparison to a more extensive alphasyllabary is ease of use with modern technology such as computers and cellphones. However, contained Latin-based alphabets still present additional technological challenges, such as in the usage of punctuation as tone marks or other graphematic diacritics. Muak Sa-aak also has reduced syllables, common throughout the region; these syllables occur before a main syllable and have limited phonological contrast in that they do not distinguish vowel quality, length, phonation, or tone. Many Latin-based orthographies in the area, such as Akha or Lahu, are written with syllable spacing rather than word spacing; alphasyllabaries such as Thai or Burmese, by contrast, tend to use phrasal spacing.

Any minority language orthography development today must take into account the Unicode standard for encoding of characters. Computers store characters such as letters, punctuation, or other symbols as numbers (codepoints). Prior to Unicode, there have been many different encodings in use for this conversion between the characters appearing on a screen and the underlying numerical codepoints. Orthographies could use the same codepoint for different characters, or encode identical characters with different numerical codepoints. Fonts were often designed to handle non-Latin symbols, IPA characters or diacritics to modify Latin characters by choosing an unneeded character on the keyboard and assigning a different character to its numerical codepoint. These have been called 'custom encoded fonts' or 'hacked fonts' and sometimes are referred to now as 'legacy fonts'. The result was that the same codepoint might encode different letters or symbols, depending on the font in use. This creates conflicts that can result in corruption of data. The Unicode Standard (The Unicode Consortium, 2019) deals with this problem by providing a single unified encoding system. It encodes all characters as hexadecimal numbers, allowing for a vast number of codepoints so that each character can be assigned a unique numerical codepoint. This remains constant across computer operating systems, programs, languages and fonts. Most major world language orthographies are included in Unicode already. For newly created orthographies in minority languages, compliance with the Unicode standard is an important consideration in any language developmental efforts.

2 Muak-Sa-aak Phoneme-Grapheme Correspondences

Muak Sa-aak is an isolating tonal language, with only a few derivational prefixes and no morphophonemic rules. For an in-depth phonological description see Hall (2013). Compounding is very common. Muak Sa-aak has many sesquisyllabic words with a reduced syllable lacking tonal and vowel contrast and having a reduced onset inventory preceding a main syllable, common across the languages of mainland Southeast Asia (cf. Matisoff 1989). The basic word structure may be summarized as (C).(C)(C)V(C)^{T}. The three Muak Sa-aak tones are low, high-falling, and constricted (glottal-final or laryngealized), with the latter having a high allotone in short syllables and a falling allotone in long syllables. The constricted tone occurs on both open[2] and closed syllables. Tone carries a low functional load and is closely linked to syllable structure, as not all syllable types co-occur with all three tones. There are three contrastive tones on smooth syllables, as in *tʰâm* 'egg', *tʰàm* 'from', *tʰậm* 'cave'. Pitch is predictable on stop-final syllables, where it is high with short and low with long vowels, as in lát 'go' and là꞉t 'be afraid'. This is further illustrated by the examples in Table 1.

[2] Syllables with a glottal stop final are phonologically open since the glottal stop is phonologically interpreted as tone.

Table 1: *Minimal sets for Muak Sa-aak tone (from Hall 2014). Shading denotes combinations that do not occur.*

Syllable type ➜		Smooth				Checked	
Tone	Phonetic realization	sonorant-final long vowel	sonorant-final short vowel	open long	open short	long	short
1	low, stiff	rà:ŋ ‘abandoned’	ràŋ ‘rich’	cì: ‘sap’		kʰà:p ‘chin’	
2	high modal (short)		ra̱ŋ [ra̱ŋ:ʔ] ‘fallow field’		cí [cíʔ] ‘do’		kʰáp ‘enough’
	high-falling creaky (long)	râ:ŋ [râ:ŋʔ] ‘separate’		cî: ‘point’			
3	high-falling, modal	râ:ŋ ‘flower’	râŋ ‘shining’	cù: cî: ‘make a hole’			

Initial consonant clusters are formed with stops followed by /w/ and, more often, /r/: /pr, pʰr, pw, pʰw, kr, kʰr, kw, kʰw/. Initial consonants in reduced syllables are limited to /p, pʰ, t, k, kʰ, m, s/. Only /p, t, c, k, m, n, ɲ, ŋ, l, j, w/ occur as final consonants. Short vowels /ɛ/ and /e/ are distinguished only in closed syllables. In open syllables, these vowels vary freely and are always long. In this context, native speakers represent this phoneme with the digraph <ae>, even though they generally do not like digraphs. The diphthongs /ia/ and /ua/ in Muak Sa-aak of Wan Fai, the variety the orthography is based on, correspond to the long vowels [ɛ:] and [ɔ:], respectively, in the Wan Saw variety of Muak Sa-aak. The vowels [ua] and [ɔ:] in the Wan Fai variety are allophones (although there are a few exceptions in borrowed words); [ua] occurs in closed syllables and [ɔ:] is limited to open syllables. In the phonological transcriptions, the more frequently occurring diphthong represents this phoneme, in symmetry with the diphthong /ia/. Native speakers however write the monophthong allophone in open syllables.

The 21 consonants and 18 vowels of Muak Sa-aak and their orthographic representation are shown in Tables 2 and 3. The phonemic transcriptions of Muak Sa-aak in IPA follow Hall (2013, 2014). Apart from the semivowels /w, j/, graphemes do not differ depending on their position in a word. Semivowels are spelled as consonants in initial position, and as vowels in medial or final position, i.e. /w/ is symbolized with <v> in initial, with <o> in final position, and <u> as medial consonant in clusters; /j/ is spelled with <y> in initial and with <i> in final position. Table 3 shows Muak Sa-aak vowels and their corresponding graphemes.

Table 2: *Muak Sa-aak consonant phonemes with proposed graphemes*

Consonant phonemes	Consonant Graphemes	Consonant phonemes	Consonant Graphemes
/p/	<p>	/s/	<s>
/pʰ/	<ph>	/h/	<h>
/b/	<b>	/w/	<v-> (initial) <-u-> (medial) <-o> (final)
/t/	<t>	/r/	<r>
/tʰ/	<th>	/l/	<l>
/d/	<d>	/j/	<y-> (initial) <-i-> (final)
/c/	<c>	/pr/	<pr>
/cʰ/	<ch>	/pʰr/	<phr>
/k/	<k>	/kr/	<kr>

/kʰ/	<kh>	/kʰr/	<khr>
/m/	<m>	/pw/	<pu>
/n/	<n>	/pʰw/	<phu>
/ɲ/	<ny>	/kw/	<ku>
/ŋ/	<ng>	/kʰw/	<khu>
/f/	<f>		

Table 3: *Muak Sa-aak vowel phonemes with proposed graphemes*

Vowel phonemes	Vowel graphemes
/i/	<i>
/iː/	<ii>
/e, eː/	<e>
/ɛ/	<ae>
/ɯ, ɯː/	<ee>
/ɤ, ɤː/	<eu>
/a/	<a>
/aː/	<aa>
/u, uː/	<u>
/o, oː/	<o>
/ɔ/	<aw>
/ia/	<ia>
/ua/	<ua>

Vowel notation is not consistent, as doubled letters are used to denote length for /a, i/, but denote a separate vowel quality <ee> for /ɯ/. This use of <ee> for /ɯ/ follows the conventions of a larger and influential neighboring minority language with an established Latin script, which does not have contrastive vowel length. As mentioned before, in some Muak Sa-aak villages the diphthongs /ia, ua/, symbolized with <ia, ua>, correspond to /ɛː, ɔː/.

There are very few minimal pairs for contrastive vowel length except for the open vowels /a, aː/. The speakers initially recognized a length contrast for close front vowels /i, iː/ and mid back vowels /o, oː/ as well, but were only willing to write the distinction for the vowels /a, aː/ and /i, iː/ because they were not really aware of the distinction. In a later orthography workshop, the speakers recognized the length distinction for the vowels /e, ɤ, u, ɯ/ as well and wanted to distinguish length for all of these in the orthography, which posed a problem. Their sense of aesthetics and desire to type their language easily on an ordinary cell phone forbids the use of diacritics. The remaining solutions, such as marking length with a consonant grapheme that was not yet in use (<j> or <z>) or using trigraphs to represent certain vowels, were either immediately unacceptable or judged to cause potential problems with a broader acceptance of the orthography. This resulted in the decision to under-represent the vowel inventory, with the plan to test the orthography and see whether this would cause a problem. Since the functional load of vowel length is low, and there are few words that are actually ambiguous, it was hoped that well-constructed literacy materials that teach reading in context and better teaching methods would obviate the need to add five more vowel graphemes to the orthography.

Writing reduced syllables is challenging because they lack contrastive tone or vowel quality. Orthographies in this region must take this phenomenon into account, seen in the following. Representing Muak Sa-aak minor syllable vocoids with a major syllable vowel symbol implies that the syllable is a full syllable. Therefore in the orthography workshop, this first was avoided through distinguishing minor syllable vowel graphemes from major syllable vowel graphemes. Participants attempted to mark minor syllables by writing only the consonantal onset, attached to the following major syllable with no space as in <smox> for /s.mox/ ‘stone’ as found in the Burmese alphasyllabary. However, there are a handful of sesquisyllables (one-and-a-half-syllabic words, cf. Matisoff 1989) where the major syllable begins with a vowel, resulting in ambiguity. One example of this is the name of the language: <muak saak> could have been pronounced as [muak sə̆ʔaːk] or [muak saːk]. A hyphen was then chosen to mark the break between minor and major syllables; this way, the language name was written as <muak s-aak> for /muak s.aːk/. This preserved a

distinction between the reduced syllable and the major syllable. In a later workshop, it was decided that this was too confusing because speakers found it too hard to distinguish between reduced syllables and some of the grammatical prefixes. Speakers were frequently unsure whether to use the hyphen or the vowel <ee> when writing. As a result the hyphen was finally removed from the orthography, and <ee> used for both syllable types. Therefore, the name of the language is now written as <muak see aak>.

2.1 Marking of Tone

Tone has a low functional load, and several speakers had initially expressed a desire not to mark it. In a follow-up workshop, however, they chose to mark the High Tone with the colon <:>, mirroring the written Burmese High Tone <◌ း>. The Constricted Tone is marked with a final <-x> in short open syllables, where it originally was meant to symbolize the phonetic glottal stop that frequently occurs in short open syllables. It is also marked with <-x> in sonorant final syllables and with a final <-q> in long open syllables. The latter syllable type only occurs with this one tone; speakers did not see the allotonic variation but rather thought of marking the accompanying creaky voice. Low tone is left unmarked in the orthography.

Table 4: *Tone marking by syllable types*

High-falling tone	< : >
Low tone	unmarked
Constricted tone: High allotone in short syllables	< -x > (open syllables) unmarked (stop final, short vowel)
Constricted tone: Falling allotone in long syllables	< -x > (sonorant final) < -q > (long creaky vowels)

There is a potential for ambiguity between the low tone words having stop final consonants, and constricted tone words with stop final consonants, since length is not marked for all vowels. For example, the orthographic representation <lek> could represent /leːk¹/ or / lek²/. In reading lists of words or syllables with no context, this ambiguity causes confusion.

2.2 Punctuation and capitalization

The speakers decided to mark the ends of sentences with a period. Commas are being used occasionally to mark a pause, and some speakers use question marks sentence-finally to mark questions. Spaces are being used between each syllable, including reduced syllables. They have not yet begun using punctuation to mark direct speech. Speakers decided not to use any capital letters in the orthography. The rationale given was that for those who did not know their English letters already, capitalization would require learning more letters.

3. Evaluation

Muak Sa-aak orthography generally reflects the phonology, with some exceptions. Vowel length is underspecified, with length contrast being written for only some vowels, and the use of doubled vowels <aa, ee, ii> is inconsistent, with two of these digraphs indicating length difference and one indicating a different vowel quality. The grapheme <ee> in particular may represent three different vowel lengths: the short vowel /ɯ/, the long vowel /ɯː/, or the very short predictable vocoid found in reduced syllables. Tone is underspecified, especially in distinguishing between the low tone and some constricted tone words. The glottal stop, although phonologically a feature of tone, is treated orthographically as a consonant. There is a one-to-two mapping for semivowels, depending on where in the syllable they occur.

As speakers began to practice writing, one of the greatest difficulties was in choosing the correct final stop, possibly because all involved speakers were educated and literate in Burmese, which does not have final stops. When they were taught to stop and think about the position of the tongue, and associate each final stop with a specific position, it seemed to help them to spell correctly. Among consonant graphemes, the alveolopalatal final stop was particularly challenging, since, although speakers recognized it as different from the alveolar or velar finals, they did not initially recognize it as identical with the initial alveolopalatal stop, which is affricated in this context. For example, [vaːⁱc] ‘sword’ is currently rendered as <vaac>. An

alternative spelling option would be writing the allophonic [i] offglide followed by an alveolar nasal or stop symbol: <Vin, Vit>. The word [vaːⁱc] 'sword' would then be written as <vaait>. Testing this alternative is an area for further study.

Both tone and vowel length are underspecified. The underspecification of vowel length is the result of speakers initially not recognizing that the vowel length contrast extended to more than three vowel qualities. When they became willing[3] to write length contrast for all seven vowel qualities with contrastive length, it was decided that fully differentiating length in the orthography would require too many digraphs; it was hoped that in context, this would not create a problem. This highlights a difficulty of Latin-based orthographies for Southeast Asian languages, which frequently have extensive vowel inventories and may have tone as well. Either some features will be underspecified, or diacritics and/or digraphs become necessary. In the Muak Sa-aak orthography, a combination of digraphs and a certain degree of underspecification have been accepted. It remains to be tested how reader-friendly this orthography is.

Tonal underspecification results mainly from the original desire of speakers not to mark tone at all. Since tone carries a low functional load, this does not seem to be problematic. The one tone that is clearly and consistently marked with <:> presents a particular technical issue relating to Unicode. This and other issues of text engineering will be discussed in the following sections.

4. Technological considerations

The orthographic choices for Muak Sa-aak have many advantages technologically, including the choice to avoid special diacritics. Although the contained Latin alphabet poses many challenges in a language with an extensive sound system of 21 consonants, 18 vowels and three tones, it does allow for immediate use of any computer for email, or any cellphone for text messaging. While keyboarding options are available for minority language orthographies, the need for installation of special keyboards or fonts creates an obstacle to use of the orthography with computers and cellphones. This in turn could discourage the use of the mother tongue in this important and popular context. In this case study, the orthographic choices ensured the immediate use of the mother tongue for text messaging at a time when they were not yet able to easily type any of the surrounding languages of wider communication on a cell phone. The widespread adoption of the Unicode standard, however, carries implications for minority orthographies, including those using Latin scripts.

In addition to unique codepoints mentioned in the introduction, specific character properties are assigned to all unicode characters so that computers recognize how to handle a character based upon these properties. Letter, Mark, Number, Punctuation, Symbol, and Separator are all character properties. They are important because they determine such matters as valid places for line-breaks or word-breaks in text. Some punctuation marks, for example, are recognized as valid for line-breaking of text. Some major classes include general, case, and normalization properties. Some glyphs may be identical in appearance but be encoded differently, with different character properties.

Useful recommendations are provided in Unicode Technical Note #19 for development of new orthographies (Anderson, McGowan, Whistler and Priest 2005). Unicode characters are arranged in blocks of similar characters; it is recommended that for Latin script orthographies, only characters from the Latin blocks should be used; for example, characters should not be used from Chinese or Arabic. IPA characters may also be used in Latin script orthographies, although these characters do not necessarily have uppercase versions. If characters from the same block are used, they are more likely to display well with available fonts. Unicode character properties, such as letter, punctuation, or number, should also be followed as computers handle the text according to these properties. If speakers still want to use punctuation to mark tone, SIL (2018) gives a guide to the potential difficulties for each of these choices. Although speakers of a minority language are free to use any orthography they wish, it will be easier from the text-engineering point of view if they use only characters that are already incorporated in Unicode. Characters from widely used scripts are more likely to be supported by software and fonts.

[3] Orthography development for Muak Sa-aak as described by Page (2013) is not purely participatory but follows a facilitated participatory approach where a considerable amount of input and help is given, see Lew (2019).

5. Application to Muak Sa-aak orthography

The orthographic choices made in the case of Muak Sa-aak are technologically more advantageous than a non-Latin script or a modified Latin script. They do not require installation of special fonts or keyboards for use, which would be an extra barrier to the use of the orthography, especially if the users are not technically savvy enough to do this themselves. However, these orthographic choices do create challenges through the use of punctuation marks as letters. The primary way in which the Muak Sa-aak orthography breaks the text-engineering rules is the use of a punctuation mark, the colon, as a diacritic to mark tone; the earlier orthography draft also used another punctuation mark, the hyphen, to mark the phonemically vowel-less minor syllables. In Unicode, punctuation marks have different character properties than letters. Therefore, the use of a symbol or a numeral as a letter may create unexpected complications.

The Muak Sa-aak orthography uses a colon with Unicode codepoint 003A to mark tone. According to the Unicode Standard (see Unicode Annex 14 and 29, the Unicode Consortium) this character does not permit line-breaking or word-breaking before or after it. Therefore this tends not to cause a problem, even if it were being used medially in multisyllabic words. It does still create some difficulties; for example, the colon is not valid in computer filenames. The hyphen originally used in the Muak Sa-aak orthography, however, is different. The codepoint 002D or hyphen minus, as found on an English keyboard layout does permit following line-breaking. For example, the Muak Sa-aak word <s-mox> 'stone' may be broken between <s-> and <mox>, leaving only 's-' hanging at the end of a line.

An alternative for the hyphen, which is a punctuation mark, or the use of the main syllable vowel <ee>, which creates ambiguity, might have been to use the codepoint 02D7 'modifier letter minus sign' instead. The 'modifier letter minus sign', 02D7, is classed as a letter (AL) for line-breaking and as a mid-letter (ML) for word-breaking; this prevents line-breaking both before and after the hyphen. Similarly, A789 'modifier letter colon' could be used in place of 003A, 'colon,' since that would be handled as a letter. However, neither of these look-alike characters, 02D7 and A789, appear on the standard computer keyboard for English or other majority languages, and would require the use of a special keyboarding solution. In practice it would probably not be used; most people would simply use the hyphen-minus, 002D and colon, 003A.

Another option is that segmentation rules for a specific language may be written so that a character may be handled in a way other than that specified in Unicode Annex 14 and 29. If the text is tagged for the language in use, and if the software used is able to implement this, text may be broken in the correct places. However, in practice most software does not yet implement this.

6. Summary and conclusion

The process of orthography development for Muak Sa-aak followed what can be labeled as a facilitated participatory approach (Lew 2019), combining phonological accuracy and the community's preferences in the orthography. Upon the speakers' request the orthography is based on Latin script, with no diacritics or special characters other than a colon marking a High Tone like in Burmese script. This led to some underspecification of vowels and tone, with the eighteen phonemic vowels being written with only thirteen graphemes, and only one of three tones consistently marked.

The Muak Sa-aak have had little access to computers but modern technology will come. They already have cell phones. The orthographic choices have technological disadvantages and advantages. The lack of diacritics or special characters not found on a standard computer keyboard allows the language to be typed easily, but the use of punctuation marks as a letters causes text processing issues. Many of the predictive keyboards used on cellphones easily adapt to the spelling of new words. However, punctuation is not recognized as letters, which results in keyboards not inserting the full word. The use of a hyphen originally marking minor syllables sometimes resulted in separating the minor syllable onset from the following major syllable. It was eventually eliminated because speakers found it too hard to distinguish between reduced syllables and certain grammatical prefixes. The colon used as a tone mark was treated as punctuation by default in Fieldworks Language Explorer (SIL); it had to be manually changed in order to be treated as a word-forming character. Similar problems could be expected to occur in other programs with punctuation marks that are used as letters, but it may not always be possible to force the software to treat a punctuation mark as a letter. An additional challenge was that the colon is not permitted in computer file names. This created confusion for speakers learning to use the computer. They did not understand why some of their file names were not accepted by the computer, further complicated by the English user interface because the

warning messages generated were incomprehensible to them. Current practice is to either ignore the tone mark or to replace it with the semicolon, which is permitted in file names.

The decision not to use any capital letters causes obstacles in the use of word processing software such as Microsoft Word, and with smartphone keyboards. The autocorrect function by default automatically capitalizes any letter following a period, recognizing the beginning of a new sentence. This setting may be changed, but in practice, speakers who are new computer users may not know how to do that. This is the first time the Muak Sa-aak people have tried to write their language, and it remains to be seen how much they will use and revise it. Several changes have already been made. A tone mark has been added, and spelling of reduced syllables has been changed twice. Their refusal to use diacritics circumvents the need for special keyboarding solutions, which makes the use of their own language in these areas more likely.

References

The Unicode Consortium. 2019. The Unicode Standard, Version 12.1.0, (Mountain View, CA: The Unicode Consortium. ISBN 978-1-936213-25-2) http://www.unicode.org/versions/Unicode12.1.0/

Anderson, Deborah, Rick McGowan, Ken Whistler and Lorna Priest. 2005. Unicode Technical Note #19, "Recommendations for Creating New Orthographies," http://www.unicode.org/notes/tn19/tn19-1.html

Heninger, Andy, ed. 2019. Unicode Standard Annex #14: "Unicode line-breaking algorithm," an integral part of The Unicode Standard. Version 12.1.0. 2019-02-15. (http://www.unicode.org/reports/tr14/tr14-43.html)

Davis, Mark ed. 2019. Unicode Standard Annex #29: "Unicode text segmentation," an integral part of The Unicode Standard. Version 12.1.0. 2019-02-15. (http://www.unicode.org/reports/tr29/tr29-35.html)

Hall, Elizabeth. 2013. A phonological description of Muak Sa-aak. *Mon-Khmer Studies Journal,* vol. 42, pp. 26-39.

Hall, Elizabeth. 2014. An analysis of Muak Sa-aak tone. *Journal of the Southeast Asian Linguistics Society*, 7:1-10.

Hosken, Martin. 2003. Creating an orthography description. http://scripts.sil.org/cms/scripts/render_download.php?format=file&media_id=OrthographyDescPaper&filename=OrthographyDescPaper.pdf retrieved June 13, 2016.

Inglis, Douglas M. 2017. Myanmar-based Khamti Shan orthography. *Journal of the Southeast Asian Linguistics Society* 10.1, xlvii-lxi. https://evols.library.manoa.hawaii.edu/bitstream/10524/52404/1/11Inglis2017myanmar.pdf

Lew, Sigrid. 2019. Proposing a facilitated participatory approach for Southeast Asian minority language orthography design. *Journal of the Southeast Asian Linguistics Society 12*(1), lxvii-lxxv. http://hdl.handle.net/10524/52451

Matisoff, J.A. 1989. The bulging monosyllable, or the mora the merrier: echo-vowel adverbialization in Lahu, in *South-east Asian Linguisitics: Essays in Honour of Eugénie J.A. Henderson*, ed. J.H.C.S. Davidson, pp. 163-198. School of Oriental and African Studies, University of London.

Nag, Sonali. 2007. Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading*, 30(1), 7-22.

Owen, Robert Wyn, 2017. A description and linguistic analysis of the Tai Khuen writing system. *Journal of the Southeast Asian Linguisitcs Society* 10.1, 140-164.

Page, Christina Joy. 2013. A new orthography in an unfamiliar script: A case study in participatory engagement strategies. *Journal of Multilingual and Multicultural Development*, pp. 1-16.

SIL. 2018. Best practice when using non-alphabetic characters in orthographies: Helping languages succeed in the modern world. https://www.sil.org/orthography/fonts-and-technical-issues Accessed 8 June 2019.

Reviewed: Received 13 June 2019, revised text accepted 10 July 2019, published DAY MONTH YEAR
Editors: Editor-In-Chief Dr Mark Alves | Managing Eds. Dr Paul Sidwell, Dr Nathan Hill, Dr Sigrid Lew